



ENTERED

1440

1500

1600

RAW SEQUENCE LISTING

DATE: 03/11/2003

PATENT APPLICATION: US/09/677,584

TIME: 10:54:53

Input Set : A:\PTO.VSK.txt

Output Set: N:\CRF4\03112003\I677584.raw

```
2 <110> APPLICANT: DIVERSA CORPORATION
3 SHORT, Jay

W--> 4 <120> TITLE OF INVENTION: WHOLE CELL ENGINEERING BY MUTAGENIZING A SUBSTANTIAL PORTION

OF A

W--> 5 STARTING GENOME, COMBINING MUTATIONS, AND OPTIONALLY REPEATING

7 <130> FILE REFERENCE: DIVER1510WO-1

C--> 9 <140> CURRENT APPLICATION NUMBER: US/09/677,584

C--> 10 <141> CURRENT FILING DATE: 2000-09-30

12 <150> PRIOR APPLICATION NUMBER: US 09/677,584

13 <151> PRIOR FILING DATE: 2000-09-30
```

15 <150> PRIOR APPLICATION NUMBER: US 09/594,459

16 <151> PRIOR FILING DATE: 2000-06-14

18 <160> NUMBER OF SEQ ID NOS: 33

20 <170> SOFTWARE: PatentIn version 3.0

22 <210> SEQ ID NO: 1 23 <211> LENGTH: 5818 24 <212> TYPE: DNA

25 <213> ORGANISM: Escherichia coli

27 <400> SEQUENCE: 1

60 28 ctgactetta tacacaagta gegteetgaa eggaacettt eeegttttee aggatetgae 120 30 ttccatgtga cctcctaaca tggtaacgtt catgataact tctgctcttc atcgtgcggc 180 32 cgactgggct aaatctgtgt tetettegge ggegetgggt gateetegee gtactgeeeg 34 cttggttaac gtcgccgccc aattggcaaa atattctggt aaatcaataa ccatctcatc 240 36 agagggtagt gaagccatgc aggaaggcgc ttaccgattt taccgcaatc ccaacgtttc 300 38 tgccgaggcg atcagaaagg ctggcgccat gcaaacagtc aagttggctc aggagtttcc 360 40 cgaactgctg gccattgagg acaccacctc tttgagttat cgccaccagg tcgccgaaga 420 42 gcttggcaag ctgggctcta ttcaggataa atcccgcgga tggtgggttc actccgttct 480 44 cttgctcgag gccaccacat tccgcaccgt aggattactg catcaggagt ggtggatgcg 540 600 46 cccggatgac cctgccgatg cggatgaaaa ggagagtggc aaatggctgg cagcggccgc 48 aactagccgg ttacgcatgg gcagcatgat gagcaacgtg attgcggtct gtgaccgcga 660 50 agccgatatt catgcttatc tgcaggacag.gctggcgcat aacgagcgct tcgtggtgcg 720 52 ctccaagcac ccacgcaagg acgtagagtc tgggttgtat ctgatcgacc atctgaagaa 780 54 ccaaccggag ttgggtggct atcagatcag cattccgcaa aagggcgtgg tggataaacg 840 900 56 cggtaaacgt aaaaatcgac cagcccgcaa ggcgagcttg agcctgcgca gtgggcgcat 960 58 cacgctaaaa caggggaata tcacgctcaa cgcggtgctg gccgaggaga ttaacccgcc 1020 60 caagggtgag accccgttga aatggttgtt gctgaccggc gaaccggtcg agtcgctagc 62 ccaageettg egegteateg acatttatae ccategetgg eggategagg agttecataa 1080 64 ggcatggaaa accggagcag gagccgagag gcaacgcatg gaggagccgg ataatctgga 1140 66 gcggatggtc tcgatcctct cgtttgttgc ggtcaggctg ttacagctca gagaaagctt 1200 68 cacgctgccg caagcactca gggcgcaagg gctgctaaag gaagcggaac acgtagaaag 1260 1320 70 ccagtccgca gaaacggtgc tgaccccgga tgaatgtcag ctactgggct atctggacaa 1380 72 gggaaaacgc aagcgcaaag agaaagcagg tagcttgcag tgggcttaca tggcgatagc

74 tagactgggc ggttttatgg acagcaagcg aaccggaatt gccagctggg gcgccctctg 76 gtaaggttgg gaagccctgc aaagtaaact ggatggcttt cttgccgcca aggatctgat

Input Set : A:\PTO.VSK.txt

78	ggcgcagggg	atcaagatct	gatcaagaga	caggatgagg	atcgtttcgc	atgattgaac	1560
			tctccggccg				1620
82	gggcacaaca	gacaatcggc	tgctctgatg	ccgccgtgtt	ccggctgtca	gcgcaggggc	1680
84	gcccggttct	ttttgtcaag	accgacctgt	ccggtgccct	gaatgaactg	caggacgagg	1740
			gccacgacgg				1800
88	tcactgaagc	gggaagggac	tggctgctat	tgggcgaagt	gccggggcag	gatctcctgt	1860
90	catctcacct	tgctcctgcc	gagaaagtat	ccatcatggc	tgatgcaatg	cggcggctgc	1920
92	atacgcttga	tccggctacc	tgcccattcg	accaccaagc	gaaacatcgc	atcgagcgag	1980
94	cacgtactcg	gatggaagcc	ggtcttgtcg	atcaggatga	tctggacgaa	gagcatcagg	2040
96	ggctcgcgcc	agccgaactg	ttcgccaggc	tcaaggcgcg	catgcccgac	ggcgaggatc	2100
98	tcgtcgtgac	ccatggcgat	gcctgcttgc	cgaatatcat	ggtggaaaat	ggccgctttt	2160
100	ctggattcat	cgactgtggc	cggctgggtg	tggcggaccg	ctatcaggac	atagcgttgg	2220
			gagcttggcg				2280
			tcgcagcgca				2340
			tcgaaatgac				2400
			ccttctatga				2460
			agcgcgggga				2520
			ggttcagctg				2580
			teggeateca				2640
			ggggaggcac				2700
			acgaattgct				2760
		-	gtggatggag				2820
			gatacttccc				2880
			tgaagaggcc				2940
			cgcctcggcg				3000
			gatagcgtgg				3060
			tetgeacgeg				3120
			geegettite				3180
			ccaacttctg				3240
			cgtgctcgac				3300
			gegeacette				3360
			gcttgcgatc				3420
			gtttgagccc				3480
			gttcatcggc				3540
			catggcacgc				3600
			tcaggccgtg				3660
			gccctgcacc				3720
			gatctgcccg				3780
			ctgaaatcgc				3840
			cagcagttga				3900
			taggcgatac				3960
			cagcgcgtga				4020
			cgcgaaggac				.4080
			cgttgccggc				4140
			tgacgccaag				4200
			gctcgccgcg				4260
			gcccctgtc				4320
			agccatccag				4380
							4440
T / 4	gayyycyccc	. cayctyycaa	ttccggttcg	citycigidd	acadactyc	ccagictage	7770

Input Set : A:\PTO.VSK.txt

```
176 tategecatg taageceact geaagetace tgetttetet ttgegettge gtttteett
                                                                         4500
178 gtccagatag cccagtagct gacattcatc cggggtcagc accgtttctg cggactggct
                                                                         4560
                                                                         4620
180 ttctacgtgt tccgcttcct ttagcagccc ttgcgccctg agtgcttgcg gcagcgtgaa
                                                                         4680
182 gctttctctg agctgtaaca gcctgaccgc aacaaacgag aggatcgaga ccatccgctc
184 cagattatec ggeteeteea tgegttgeet eteggeteet geteeggttt tecatgeett
                                                                         4740
186 atggaactcc tcgatccgcc agcgatgggt ataaatgtcg atgacgcgca aggcttgggc
                                                                         4800
188 tagcgactcg accggttcgc cggtcagcaa caaccatttc aacggggtct cacccttggg
                                                                         4860
190 cqqqttaatc tcctcqqcca qcaccqcqtt qagcqtqata ttcccctqtt ttaqcqtqat
                                                                         4920
                                                                         4980
192 gegeecactg egeaggetea agetegeett gegggetggt egatttttae gtttacegeg
                                                                         5040
194 tttatccacc acgccctttt gcggaatgct gatctgatag ccacccaact ccggttggtt
196 cttcaqatqq tcqatcaqat acaacccaqa ctctacqtcc ttqcqtqqqt gcttqqagcq
                                                                         5100
198 caccacqaaq cqctcqttat gcqccaqcct gtcctqcaqa taaqcatqaa tatcqqcttc
                                                                         5160
200 geggteacag acegeaatea egttgeteat catgetgeee atgegtaace ggetagttge
                                                                         5220
202 ggccgctgcc agccatttgc cactctcctt ttcatccgca tcggcagggt catccgggcg
                                                                         5280
204 catccaccac teetgatgea gtaateetae ggtgeggaat gtggtggeet egageaagag
                                                                         5340
206 aacggagtga acceaceate egegggattt ateetgaata gageeeaget tgeeaagete
                                                                         5400
208 ttcggcgacc tggtggcgat aactcaaaga ggtggtgtcc tcaatggcca gcagttcggg
                                                                         5460
210 aaactcctga gccaacttga ctgtttgcat ggcgccagcc tttctgatcg cctcggcaga
                                                                         5520
212 aacqttggga ttgcggtaaa atcggtaagc gccttcctgc atggcttcac taccctctga
                                                                         5580
214 tgagatggtt attgatttac cagaatattt tgccaattgg gcggcgacgt taaccaagcg
                                                                         5640
216 ggcagtacgg cgaggatcac ccagcgccgc cgaagagaac acagatttag cccagtcggc
                                                                         5700
218 cgcacgatga agagcagaag ttatcatgaa cgttaccatg ttaggaggtc acatggaagt
                                                                         5760
220 cagateetgg aaaacgggaa aggtteegtt caggacgeta ettgtgtata agagteag
                                                                         5818
223 <210> SEQ ID NO: 2
224 <211> LENGTH: 476
225 <212> TYPE: PRT
226 <213> ORGANISM: Escherichia coli
228 <400> SEQUENCE: 2
230 Met Ile Thr Ser Ala Leu His Arg Ala Ala Asp Trp Ala Lys Ser Val
231 1
233 Phe Ser Ser Ala Ala Leu Gly Asp Pro Arg Arg Thr Ala Arg Leu Val
                                    25
236 Asn Val Ala Ala Gln Leu Ala Lys Tyr Ser Gly Lys Ser Ile Thr Ile
239 Ser Ser Glu Gly Ser Glu Ala Met Gln Glu Gly Ala Tyr Arg Phe Tyr
240
242 Arg Asn Pro Asn Val Ser Ala Glu Ala Ile Arg Lys Ala Gly Ala Met
243 65
                        70
245 Gln Thr Val Lys Leu Ala Gln Glu Phe Pro Glu Leu Leu Ala Ile Glu
246
                    85
                                        90
248 Asp Thr Thr Ser Leu Ser Tyr Arg His Gln Val Ala Glu Glu Leu Gly
               100
                                    105
251 Lys Leu Gly Ser Ile Gln Asp Lys Ser Arg Gly Trp Trp Val His Ser
            115
                                120
254 Val Leu Leu Glu Ala Thr Thr Phe Arg Thr Val Gly Leu Leu His
                            135
                                                140
        130
257 Gln Glu Trp Trp Met Arg Pro Asp Asp Pro Ala Asp Ala Asp Glu Lys
260 Glu Ser Gly Lys Trp Leu Ala Ala Ala Ala Thr Ser Arg Leu Arg Met
```

Input Set : A:\PTO.VSK.txt

```
165
263 Gly Ser Met Met Ser Asn Val Ile Ala Val Cys Asp Arg Glu Ala Asp
               180
                                   185
266 Ile His Ala Tyr Leu Gln Asp Arg Leu Ala His Asn Glu Arg Phe Val
                               200
269 Val Arg Ser Lys His Pro Arg Lys Asp Val Glu Ser Gly Leu Tyr Leu
                            215
272 Ile Asp His Leu Lys Asn Gln Pro Glu Leu Gly Gly Tyr Gln Ile Ser
                        230
                                            235
275 Ile Pro Gln Lys Gly Val Val Asp Lys Arg Gly Lys Arg Lys Asn Arg
                   245
                                       250
278 Pro Ala Arg Lys Ala Ser Leu Ser Leu Arg Ser Gly Arg Ile Thr Leu
               260
                                    265
281 Lys Gln Gly Asn Ile Thr Leu Asn Ala Val Leu Ala Glu Glu Ile Asn
           275
                                280
284 Pro Pro Lys Gly Glu Thr Pro Leu Lys Trp Leu Leu Thr Gly Glu
                           295
287 Pro Val Glu Ser Leu Ala Gln Ala Leu Arg Val Ile Asp Ile Tyr Thr
                                            315
290 His Arg Trp Arg Ile Glu Glu Phe His Lys Ala Trp Lys Thr Gly Ala
                    325
                                        330
293 Gly Ala Glu Arg Gln Arg Met Glu Glu Pro Asp Asn Leu Glu Arg Met
                340
                                    345
296 Val Ser Ile Leu Ser Phe Val Ala Val Arg Leu Leu Gln Leu Arg Glu
                                360
299 Ser Phe Thr Leu Pro Gln Ala Leu Arg Ala Gln Gly Leu Leu Lys Glu
                            375
302 Ala Glu His Val Glu Ser Gln Ser Ala Glu Thr Val Leu Thr Pro Asp
                        390
                                            395
305 Glu Cys Gln Leu Leu Gly Tyr Leu Asp Lys Gly Lys Arg Lys Arg Lys
                   405
                                        410
308 Glu Lys Ala Gly Ser Leu Gln Trp Ala Tyr Met Ala Ile Ala Arg Leu
               420
                                   425
311 Gly Gly Phe Met Asp Ser Lys Arg Thr Gly Ile Ala Ser Trp Gly Ala
                                440
314 Leu Trp Glu Gly Trp Glu Ala Leu Gln Ser Lys Leu Asp Gly Phe Leu
                            455
317 Ala Ala Lys Asp Leu Met Ala Gln Gly Ile Lys Ile
                        470
320 <210> SEQ ID NO: 3
321 <211> LENGTH: 30
322 <212> TYPE: DNA
323 <213> ORGANISM: Artificial sequence
325 <220> FEATURE:
326 <223> OTHER INFORMATION: Defined sequence kernel
328 <220> FEATURE:
329 <221> NAME/KEY: misc feature
330 <222> LOCATION: (1)..(30)
331 <223> OTHER INFORMATION: n is A, T, G, or C
```

Input Set : A:\PTO.VSK.txt

```
333 <400> SEQUENCE: 3
W--> 334 nnknnknnkn nknnknnknn knnknnknnk
                                                                                 30
     337 <210> SEQ ID NO: 4
     338 <211> LENGTH: 30
     339 <212> TYPE: DNA
     340 <213> ORGANISM: Artificial sequence
     342 <220> FEATURE:
     343 <223> OTHER INFORMATION: Defined sequence kernel
     345 <220> FEATURE:
     346 <221> NAME/KEY: misc feature
     347 <222> LOCATION: (1)..(30)
     348 <223> OTHER INFORMATION: n is A, T, G, or C
     350 <400> SEQUENCE: 4
                                                                                30
W--> 351 nnmnnmnnmn nmnnmnnmnn mnnmnnmnnm
     354 <210> SEQ ID NO: 5
     355 <211> LENGTH: 5
     356 <212> TYPE: PRT
     357 <213> ORGANISM: Artificial sequence
     359 <220> FEATURE:
     360 <223> OTHER INFORMATION: Antibody spacer peptide. The entire peptide sequence can be
               repeated more than one time
     363 <400> SEQUENCE: 5
     365 Gly Gly Gly Ser
     366 1
     368 <210> SEQ ID NO: 6
     369 <211> LENGTH: 14
     370 <212> TYPE: DNA
     371 <213> ORGANISM: Artificial sequence
     373 <220> FEATURE:
     374 <223> OTHER INFORMATION: Tetradecanucleotide d
     376 <400> SEQUENCE: 6
                                                                                 14
     377 catgccatgg catg
     380 <210> SEQ ID NO: 7
     381 <211> LENGTH: 21
     382 <212> TYPE: DNA
     383 <213> ORGANISM: Artificial sequence
     385 <220> FEATURE:
     386 <223> OTHER INFORMATION: 21-mer d
     388 <400> SEQUENCE: 7
                                                                                 21
     389 aaattgtgca catcctgcag c
     392 <210> SEQ ID NO: 8
     393 <211> LENGTH: 12
     394 <212> TYPE: DNA
     395 <213> ORGANISM: Artificial sequence
     397 <220> FEATURE:
     398 <223> OTHER INFORMATION: 12-mer target DNA
     400 <400> SEQUENCE: 8
                                                                                 12
     401 agcctagctg aa
     404 <210> SEQ ID NO: 9
```

Input Set : A:\PTO.VSK.txt

Output Set: N:\CRF4\03112003\I677584.raw

Please Note:

Use of n and/or Xaa have been detected in the Sequence Listing. Please review the Sequence Listing to ensure that a corresponding explanation is presented in the <220> to <223> fields of each sequence which presents at least one n or Xaa.

```
Seq#:3; N Pos. 1,2,4,5,7,8,10,11,13,14,16,17,19,20,22,23,25,26,28,29
Seq#:4; N Pos. 1,2,4,5,7,8,10,11,13,14,16,17,19,20,22,23,25,26,28,29
Seq#:10; Xaa Pos. 3
Seq#:12; Xaa Pos. 2
Seq#:13; N Pos. 6,7,8,9,10
Seq#:14; N Pos. 1,2,3,4,5,6,7,8,9,10,21,22,23,24,25,26,27,28,29,30,31,32,33
Seq#:14; N Pos. 34,35,36,37,38,39,40,41,42,43,44,45,46,47,48,49,50,51,52,53
Seq#:14; N Pos. 54,55,56,57,58,59,60,61,62,63,64,65,66,67,68,69,70,71,72,73
Seq#:14; N Pos. 74,75,76,77,78,79,80,81,82,83,84,85,86,87,88,89,90,91,92,93
Seq#:14; N Pos. 94,95,96,97,98,99,100,101,102,103,104,105,106,107,108,109
Seq#:14; N Pos. 110,111,112,113,114,115,116,117,118,119,120,124,125,126,127
Seq#:14; N Pos. 128,129,130,131,132,133,134,135,136,137,138,139,140,141,142
Seq#:14; N Pos. 143,144,145,146,147,148,149,150,151,152,153,154,155,156,157
Seq#:14; N Pos. 158,159,160,161,162,163,164,165,166,167,168,169,170,171,172
Seq#:14; N Pos. 173,174,175,176,177,178,179,180,181,182,183,184,185,186,187
Seq#:14; N Pos. 188,189,190,191,192,193,194,195,196,197,198,199,200,201,202
Seq#:14; N Pos. 203,204,205,206,207,208,209,210,211,212,213,214,215,216,217
Seq#:14; N Pos. 218,219,220,221,222,223
Seq#:15; N Pos. 1,2,3,4,5,6,7,8,9,10,16,17,18,19,20,21,22,23,24,25,26,27,28
Seq#:15; N Pos. 29,30,31,32,33,34,35,36,37,38,39,40,41,42,43,44,45,46,47,48
Seq#:15; N Pos. 49,50,51,52,53,54,55,56,57,58,59,60,61,62,63,64,65,66,67,68
Seq#:15; N Pos. 69,70,71,72,73,74,75,76,77,78,79,80,81,82,83,84,85,86,87,88
Seq#:15; N Pos. 89,90,91,92,93,94,95,96,97,98,99,100,101,102,103,104,105
Seq#:15; N Pos. 106,107,108,109,110,111,112,113,114,115,116,117,118,119,120
Seq#:15; N Pos. 121,122,123,124,125,126,127,128,129,130,131,132,133,134,135
Seq#:15; N Pos. 136,137,138,139,140,141,142,143,144,145,146,147,148,149,150
Seq#:15; N Pos. 151,152,153,154,155,156,157,158,159,160,161,162,163,164,165
Seq#:15; N Pos. 166,167,168,169,170,171,172,173,174,175,176,177,178,179,180
Seq#:15; N Pos. 181,182,183,184,185,186,187,188,189,190,191,192,193,194,195
Seq#:15; N Pos. 196,197,198,199,200,201,202,203,204,205,206,207,208,209,210
Seq#:15; N Pos. 211,212,213,214,215
Seq#:16; N Pos. 24,25,26,27,28,29,30,31,32,33,34,35,36,37,38,39,40,41,42,43
Seq#:16; N Pos. 44,45,46,47,48,49,50,51,52,53,54,55,56,57,58,59,60,61,62,63
Seq#:16; N Pos. 64,65,66,67,68,69,70,71,72,73,74,75,76,77,78,79,80,81,82,83
Seq#:16; N Pos. 84,85,86,87,88,89,90,91,92,93,94,95,96,97,98,99,100,101,102
Seq#:16; N Pos. 103,104,105,106,107,108,109,110,111,112,113,114,115,116,117
Seq#:16; N Pos. 118,119,120,121,122,123
Seq#:17; N Pos. 22,23,24,25,26,27,28,29,30,31,32,33,34,35,36,37,38,39,40,41
Seq#:17; N Pos. 42,43,44,45,46,47,48,49,50,51,52,53,54,55,56,57,58,59,60,61
Seq#:17; N Pos. 62,63,64,65,66,67,68,69,70,71,72,73,74,75,76,77,78,79,80,81
Seq#:17; N Pos. 82,83,84,85,86,87,88,89,90,91,92,93,94,95,96,97,98,99,100
Seq#:17; N Pos. 101,102,103,104,105,106,107,108,109,110,111,112,113,114,115
Seg#:17; N Pos. 116,117,118,119,120,121
```

VERIFICATION SUMMARY

DATE: 03/11/2003 TIME: 10:54:54

PATENT APPLICATION: US/09/677,584

Input Set : A:\PTO.VSK.txt

```
L:4 M:283 W: Missing Blank Line separator, <120> field identifier
L:9 M:270 C: Current Application Number differs, Replaced Current Application Number
L:10 M:271 C: Current Filing Date differs, Replaced Current Filing Date
L:334 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:3 after pos.:0
L:351 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:4 after pos.:0
L:431 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:10 after pos.:0
L:462 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:12 after pos.:0
L:479 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:13 after pos.:0
L:511 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:14 after pos.:0
M:341 Repeated in SeqNo=14
L:549 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:15 after pos.:0
M:341 Repeated in SeqNo=15
L:577 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:16 after pos.:0
M:341 Repeated in SeqNo=16
L:603 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:17 after pos.:0
M:341 Repeated in SeqNo=16
```